UC Irvine
UC Irvine Previously Published Works

Title
Reply to comment by K. Beven and P. Young on “Bayesian recursive parameter estimation for hydrologic models”

Permalink
https://escholarship.org/uc/item/5v1217j4

Journal
Water Resources Research, 39(5)

ISSN
0043-1397

Authors
Gupta, Hoshin
Thiemann, Michael
Trosset, Michael
et al.

Publication Date
2003-05-01

DOI
10.1029/2002wr001405

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Reply to comment by K. Beven and P. Young on "Bayesian recursive parameter estimation for hydrologic models"

Hoshin Gupta, Michael Thiemann,[1] Michael Trosset,[2] and Soroosh Sorooshian

Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA

Received 25 April 2002; revised 1 July 2002; accepted 1 July 2002; published 7 May 2003.

INDEX TERMS: 1860 Hydrology: Runoff and streamflow; 1894 Hydrology: Instruments and techniques; 9820 General or Miscellaneous: Techniques applicable in three or more fields; *KEYWORDS:* recursive estimation uncertainty, rainfall-flow models, GLUE, model errors

Citation: Gupta, H., M. Thiemann, M. Trosset, and S. Sorooshian, Reply to comment by K. Beven and P. Young on ''Bayesian recursive parameter estimation for hydrologic models,'' *Water Resour. Res.*, *39*(5), 1117, doi:10.1029/2002WR001405, 2003.

[1] We would like to begin by thanking our colleagues Keith Beven and Peter Young for highlighting [*Beven and Young*, 2003] (hereinafter referred to as BY) some of the inadequacies of the BARE (Bayesian recursive estimation) procedure published by us in *Water Resources Research* [*Thiemann et al.*, 2001] (hereinafter referred to as TTGS). Through their comment and this response, we hope to bring much needed attention to some very important issues in hydrologic model identification that the BARE and GLUE (generalized likelihood uncertainty estimation) methodologies attempt to address. Although our original paper had discussed many of these issues and pointed out problems with the current implementation of BARE, this response gives us another opportunity to clarify several points that may have not been made sufficiently strongly in the TTGS paper and may have led to some degree of misunderstanding. In particular, we note that the BY comments do not bring the underlying theoretical development of BARE into question but derive mainly from interpretations of the results of the numerical implementation of the BARE methodology in the context of the case studies that we presented.

[2] Let us first say that the BARE methodology is an attempt (certainly not the first) to pose the model identification problem in the context of a Bayesian framework, which allows us to ''quantify uncertainty about prediction in a natural and meaningful way, without recourse to calibration'' (section 2, second paragraph of TTGS). Further, the proposed BARE algorithm differs from conventional calibration and prediction methods in that it employs a recursive scheme for tracking the conditional probabilities associated with several competing parameter sets (models) in an online mode instead of searching for a single best solution in an off-line mode (section 5 first paragraph of TTGS). As we see later, there are several problems with the current numerical implementation of BARE that must be resolved in order for the methodology to realize its full potential. We do not believe these problems to be caused by limitations of the underlying assumptions of our (theoretical) analysis but as consequences of inadequacies in the (computational/numerical) implementation. The summary and discussion section of TTGS recognizes many of these inadequacies (''our experiences with the current version . . . suggest that there are several ways in which the implementation can be improved'') and points to the need for further research.

[3] With this preamble, we move on to discuss the various points raised by BY. We agree that the most critical issue raised by our paper is related to ''the assumption that the model structure is correct.'' We also agree with BY that ''a full acceptance of model structural error . . . requires a change in philosophy because it means that the calibration problem can no longer be considered as simply a matter of finding the parameter values of the model. . ..'' The development and proposed implementation of the BARE methodology is intended to reflect this view. The derivation therefore begins with a ''family of mathematical models'' which contains ''various sources of potential prediction error: (1) measurement error. . .; (2) parameter identification error . . .; and (3) model specification error in which it may be that for no $\theta \in \Theta$ is the model $\eta(\ |\theta)$ an accurate representation of the physical data generating process'' (TTGS, section 2.1). For purposes of illustrating the BARE methodology, each of our two case studies focused on just one model structure. Although we did not explicitly demonstrate, via the case studies, the use of a family of multiple models having different (incorrect) structures, the assumptions and derivation presented in TTGS, section 2, do allow the error function N_T in equation (27) to be computed simultaneously for any number of models.

[4] Clearly, when the concept of a ''true'' or ''correct'' model is discarded, the conventional concept (adopted by much previous research) of a ''true'' or ''correct'' parameter set is no longer appropriate. Historically, model calibration strategies have been pursued with the goal of finding an ''optimal'' parameter set. In doing so, the calibration approach (in hydrology) has implicitly ignored uncertainty, and the results are used to generate model predictions, which are reported without estimates of the underlying prediction uncertainty. The limitations of that approach lead us to the view (expressed in section 1, second paragraph of TTGS) that ''the aim of model calibration is to reduce the uncertainty in the correct choice of the parameter values (parameter uncertainty) while accounting for . . . data uncertainty . . . and model uncertainty or structural uncertainty. . ..'' In the context of a Bayesian approach, therefore, ''Calibration is the

[1]Now at Riverside Technology, Inc., Fort Collins, Colorado, USA.

[2]Now at Department of Mathematics, College of William and Mary, Williamsburg, Virginia, USA.

Copyright 2003 by the American Geophysical Union.
0043-1397/03/2002WR001405

problem of describing a plausible set of values for θ, the parameters of the models. To calibrate, we must first compute the marginal posterior density of θ . . . (and an) HPD region for this density describes a set of plausible models'' (TTGS, section 2.3).

[5] BY point out, however, that the BARE parameter estimates presented in our case studies converge to point estimates with apparently little or no uncertainty, and remark that ''this result is surely difficult to justify in any practical terms.'' In the context of the BARE methodology, there are two important aspects to this issue. The first has to do with the underlying approach inherent to Bayesian statistics. The second has to do with practical aspects of numerical implementation. In both cases, we assume that the user has selected a finite number of model families (structures with unknown but bounded parameters) from within which model identification is to be conducted. Note that, although the number of model families may be finite, the number of possible model/parameter set selections from within the restricted space of possible models is effectively infinite.

[6] Regarding the Bayesian approach, the issue has to do with how one represents uncertainty in the model/parameter space. Starting with any a priori statement of model/parameter uncertainty, an implementation of Bayes rule will, via a process of conditioning the model/parameter space on the available data, gradually and ultimately lead to the preponderance of the probability mass being assigned to a progressively smaller region of the model/parameter space. This will be true whether we use one or several families of models to describe the physical phenomenon. The process of assimilating the data via Bayes rule will, in the theoretical limit, concentrate the posterior on the set of best model structure-parameter sets that, from among the models tested, are the best able to explain/reproduce the actual physical process represented by the available input-output data (in the case that there is a single best model-parameter set within the models to be tested, the procedure will of course focus on it). All the residual uncertainty describing the ability of these selected models to simulate the physical process is then summarized in terms of the statistical properties of the residual error sequence. Note that this Bayesian framework only describes the ''relative support'' for one or more models in the model set against the other models in that set. It says little about the ''absolute support'' for those models in any general sense, and therefore provides no further guidance about model uncertainty. It should be clear therefore that the theoretical basis for BARE is reasonable and that the procedure will identify several models if there is no evidence in the data to discriminate between them given the underlying modeling assumptions (see assumptions used by TTGS, listed toward the end of this response). In the case studies presented in our paper, with only one testable model structure and a limited number of samples in the parameter space, the BARE procedure indicated that the data eventually support (in a relative sense) only a single ''best'' model from among the possibilities presented to it (however, see further comments about sampling below).

[7] The question then remains how to translate the information contained in the residual error sequence into statements about remaining model uncertainty (in the absolute sense). The BARE methodology clearly partitions the prediction uncertainty into that due to relative support for each member of the family of available models, and that due to error unexplainable by the family of available models. As the BARE algorithm recursively processes the data, a progressively smaller region of the model/parameter space emerges as being better supported by the data than other regions of the model parameter space. The remaining (residual) uncertainty is ascribed to a combination of data error and model structural error. In contrast, the GLUE approach maps the residual error uncertainty back into the model/parameter space in such a way that the prediction uncertainty brackets the observed data, with *none* of the prediction uncertainty being attributed to the inability of the family of available models to reproduce the real process. Certainly, this characteristic of GLUE is equally difficult to justify in practical terms.

[8] Let us now discuss the practical aspects of the numerical implementation of a Bayesian methodology in the form of a practical algorithm on modern digital computers, with particular reference to the mechanics of the BARE algorithm (although much of this also applies to GLUE). The practical development presented in our paper proposes that we approximate the infinite set of possible models with a finite set of models sampled from the model/parameter space; i.e., a model structure with specified parameter values constitutes one possible model. The implicit assumption here is that the finite set of sampled models is an adequate approximation of the entire set of possible models. However, this (invalid) assumption obviously does not properly account for the ''sampling-approximation error.'' We admit that the TTGS paper does not explicitly clarify this point and that BARE does not account for this sampling/approximation error in the derivation of the prediction uncertainty (and neither does GLUE).

[9] Given the existence of sampling-approximation error in the numerical implementation, when (if) the practical BARE algorithm assigns the majority (e.g. >95%) of the probability mass to a single sample model, this should not be interpreted as meaning that there is little or no residual uncertainty. Rather, it means that the Bayesian procedure has identified this sample model as the ''most likely model'' from within the finite sample of models. Since this most likely model is implicitly a representative of some particular neighborhood region of the model/parameter space, the uncertainty has clearly not been reduced to zero. The current numerical implementation of BARE does not suggest how to properly reflect this uncertainty in the computations of parameter and prediction uncertainty. This problem is not a consequence of the theoretical development of BARE, but of its numerical implementation. A simple way to deal with this problem might be to estimate, in some approximate way, the region of the model/parameter space that is ''represented'' by each model/parameter sample (e.g. using rectangular, hyper-spherical, or other geometrically approximated regions) and to use these sample uncertainty estimates as the basis for computing and reporting the residual parameter and prediction uncertainty.

[10] We did state clearly in our paper that the ''current version of the BARE procedure is limited to selecting from a fixed set of randomly specified points distributed rather coarsely throughout the feasible space,'' and that there are several ways this implementation ''can be improved'' (TTGS, section 5). We mentioned the fact of ''uneven and

insufficient density of sampling in the region of good parameter values," and suggested "progressive resampling in order to concentrate the samples in the current HPD region while terminating computations in the non-productive portions of the parameter space." We proposed procedures such as Latin-hypercube sampling "to control sample evenness." We also proposed the incorporation of a "forgetting factor" into the probability-updating rule, to prevent the model/parameter HPD from collapsing to a single point, and so as to not "overvalue" the information in the most recent data, and to relax the implicit assumption of a time-invariant best model/parameter. Strategies based on these procedures would help to reduce errors arising due to the discrete sample approximation of the non-discrete model/parameter space, and should be combined with the method mentioned in the previous paragraph.

[11] The existence of sampling-approximation error in the implementation of BARE is relevant to the BY comments regarding the two case studies presented in our paper. The first case study is a simulation example using the two-term Nash Cascade (NC) model. This example was designed as a synthetic study to illustrate the "ability of the BARE method to locate (i.e. assign) the highest probability to the region of the known true parameter values" (TTGS, introduction to section 4). It uses a "simple unit hydrograph model to show that the BARE algorithm does indeed provide consistent one-step-ahead probabilistic predictions of the true and measured streamflow values," and can "quickly and precisely locate the region of the known true parameter values, even when the data contain significant amounts of noise" (section 5, third paragraph of TTGS). The simple NC model was selected because it allows us to graphically illustrate the evolution of the HPD region in the two-dimensional parameter space. The synthetic study format allows an assessment of the prediction and parameter estimation potential of the Bayesian methodology under "the controlled conditions of no errors in model structure and known properties of the data measurement error."

[12] Three illustrative NC experiments were run (called strategies 1, 2, and 3 in TTGS). The first experiment demonstrated the ability of the algorithm to correctly identify the "correct" parameter values in the ideal case of perfect model structure (under both "no data error" and "20% output data error" conditions). This is surely a necessary test of any proposed methodology, from which neither BARE, nor GLUE (nor any other model identification strategy) should be considered exempt. Note that this case can be viewed as having "no model error," because the "known true" parameter values are contained within the set of sampled parameter sets. In the case of no data error it identifies the "true" model/parameter set in exactly three time steps. In the case of the 20% output data error, it assigns 95% of the probability to the "true" model/parameter set after approximately 75 days. The prediction uncertainty associated with this 95% parameter probability has therefore become negligible (under these idealized conditions of a discrete model set).

[13] The second and third experiments are somewhat more realistic in that although they have a correct model structure, they do not include the "true parameter set" (used to generate the test data). In this case there is model error associated with the finite set of discrete model/parameter samples. In both these cases, the BARE algorithm results in a slower but steady convergence of the HPD region into "the vicinity of the true parameter values" (TTGS, section 4.2.2, first paragraph). The convergence of the "output prediction uncertainty due to residual parameter uncertainty" in Figure 3 of TTGS to a single line is largely a consequence of not properly representing the parameter and output uncertainty due to sampling-approximation error – as described above. Refined sampling in the region of the HPD would be necessary to demonstrate that BARE gives reasonable estimates of residual parameter (and associated forecast) uncertainty. Perhaps we should have given more importance to demonstrating this in the paper. However, given the limitations of the numerical procedure, it does give reasonable "total" prediction error estimates with only 10% of the observations outside the confidence intervals (rather than the theoretical expected 5%).

[14] BY critique this example by noting that the parameters can be estimated in a much more computationally efficient manner (without Monte Carlo simulation) in which only the first two moments of the distribution (mean and covariance) are updated sequentially using, for example, the Refined Instrumental Variable (RIV) method. The RIV example is interesting for the insight it provides into how the Bayesian problem for non-linear systems might be formulated in a different way to the BARE or GLUE approaches. We certainly agree that there must necessarily remain some residual parameter and forecast uncertainty at any time due to prior uncertainty, data errors, model errors, and numerical/procedural inadequacies. The challenge is to ensure that these are all adequately represented. The RIV method is certainly a more effective and efficient way (than BARE) of computing the parameter and output uncertainties for the simple (and only mildly non-linear) NC model, where it is valid to assume that first and second order statistics can provide reasonable approximations of the mildly non-Gaussian probability distributions (although the methodology would have to be adapted if the prior uncertainty distributions or data uncertainties are strongly non-Gaussian). Where such is the case, the RIV method is an excellent (both efficient and effective) choice. However, a central point of the BARE approach is that it poses no theoretical or practical restrictions on the forms of the probability density functions or on the non-linearity of the model structures. It is not clear if the RIV approach could be applied to the SAC-SMA model or other such watershed models.

[15] BY continue their critique of the BARE method in the context of the "real data" Leaf River example illustrated in Figure 6 of TTGS. The stated purpose of this case study was to "explore the utility of the BARE method in an operational setting involving the prediction of streamflow for an "un-calibrated" watershed. . ." (TTGS, introduction to section 4) for which "data monitoring has only recently been initiated" (TTGS, section 4.3, first paragraph). In such a case hydrologists must "make the best use of the limited amounts of available data, crude parameter estimates (ranges) based on nearby watersheds, and new gauge data as they become available" (TTGS, section 4.3, first paragraph). Our purpose was not to identify a single parameter set for purposes such as "regionalization," "studying effects of catchment change," or to "compare outputs under

different input conditions.'' (Although, in principle, all of these things can just as well be done using posterior parameter distributions instead of point estimates!)

[16] The case study was conducted to test and evaluate the BARE concept (and its numerical implementation), under the simplified assumptions of a uniform prior distribution on the model/parameter space, and Gaussian output errors having heteroscedastic variance related to flow level that can be stabilized by the Box-Cox transformation. We state clearly (TTGS, section 3.2, last paragraph) that the BARE solution will ''be sensitive to the choice of the predictive model, the error model and transformation, and the data used.'' Further, we construct the approximation model/parameter set using a very coarse uniform sample consisting of only 10,000 randomly selected locations from the 13 dimensional parameter space. In presenting and interpreting the (not fully satisfactory) results, including the ''parameter-jumping'' phenomenon, we note that ''a possible cause is insufficient density of sampling in the region of good parameter values,'' and ''we should also consider the possibility that the assumed models (i.e. hydrologic model, error model, and transformation model) do not adequately represent the observed input-output process'' (TTGS, p. 2533). We are clear that the current implementation is unsatisfactory and that more work needs to be done to improve it. For example, we mention in section 3.1 (last paragraph) that ''although not explicitly handled in this paper, the error model could also be extended to autocorrelated errors for data where the error series of the used model is, in fact, not independent.''

[17] Be that as it may, we should mention that in our own previous research experience, we have found little consistent evidence to suggest that including correlation in the streamflow measurement errors by adding an error correlation noise model adds robustness to the identification of a conceptual watershed model such as the SAC-SMA. Other researchers have, however, indicated that consideration of autocorrelated errors seems to help in the case of transfer function and other kinds of input-output models. Referring to Figure 5 of our paper, we note that the model forecasts appear to be essentially unbiased until day 90 and perhaps extending until approximately day 180. This is the period in which the parameter uncertainty is still large (relative to the size of the parameter sampling region) and hence the effects of insufficient density of sampling are not being felt. However, when the size of the parameter HPD becomes so small (relative to the sample density) that the coarse parameter sampling causes it to become associated with only one sample point, the ''residuals show a tendency toward systematic positive bias'' (TTGS, p. 2534, last paragraph). The version of the BARE algorithm presented in the paper is ''limited to selecting from a fixed set of randomly specified points distributed rather coarsely through the parameter space,'' and does not have the ability of employing a search procedure to actively refine the sampling to concentrate in the region of the current HPD (TTGS, section 4.3, last paragraph). Because of this procedural sampling limitation the algorithm has selected a single parameter set, which is the best among the choices available but is unable to provide unbiased forecasts. As mentioned earlier (and in the paper), the BARE procedure needs improvement to resolve this issue. Improvements could include (1) including an autocorrelated error noise model, (2) progressively refining the model/parameter sampling density to track the evolution of the HPD region, and (3) computing and projecting the parameter uncertainty region associated with each sample point into the output space.

[18] As pointed out by BY, the current implementation of BARE is likely to be sensitive to the choice of calibration data set or data measurement errors. We did not present results to test this, although it would have been easy enough to do. However, such tests will perhaps be more meaningful when the methodological improvements mentioned above have been implemented. It is curious, however, that BY expect the model (plus error estimates) to bracket the observed discharge data. We are presenting 95% confidence intervals, and as such would expect approximately 5% of the observations to fall outside the confidence intervals.

[19] Having remarked on the deficiencies of the BARE procedure, BY continue by reflecting on the general problem of hydrologic model identification. They state that ''it is also well known that most of the difficulties associated with the calibration problem for rainfall-runoff models comes from assuming conceptual model structures a priori that are not reasonable representations of the system and are effectively overparameterized with respect to the information content of the (non-error-free) input and output data series. This is still a common practice, despite the problems of trying to find global optimal parameter sets with either single or multiple objective functions (as amply demonstrated by previous work of the Arizona group).''

[20] Here we must beg to differ with what BY seem to believe to be true. We ourselves do not subscribe to the opinion that the conceptual model structures such as the SAC-SMA are unreasonable representations of the watershed system for the purposes for which they are intended, nor do we believe that it has been demonstrated that such models are ''effectively overparameterized with respect to the information content of the input and output data series.'' It continues to be our contention [e.g., see *Gupta*, 2000] that one cannot draw strong conclusions about model inadequacy or model overparameterization as long as we continue to utilize weak procedures for extracting relevant information from the data. Our continued search for methodological improvements is based on the observation and therefore hypothesis that existing model identification procedures are inadequate in the context of conceptual watershed models and even more inadequate in the context of the emerging families of more complex watershed model representations. In particular, single criterion methods (based on statistical regression theory) that process the input-output data en masse (batch data processing methods) are unable to detect and extract important information from the data that are readily apparent and obvious to a trained hydrologist using visual examination of the observed and model simulated output hydrographs. We believe that the onus is on the hydrologic research community to ensure that our model identification tools are powerful and robust and well suited to the models at hand, rather than to draw weak conclusions regarding model validity and degree of overparameterization from weak identification methods.

[21] Finally, BY mention some recent responses to the problems of model identifiability, including (1) improving the identifiability of the chosen model, (2) the use of

"better" single and multiple performance measures, and (3) to confront the problem of model structural error directly. The GLUE methodology is represented as being a method that adopts the latter philosophical approach. It should be clear that the BARE methodology does the same. We agree that BARE and GLUE represent two different approaches, each with its own deficiencies. We have already explained (in our original paper) our view about the limitations of GLUE as a proper Bayesian approach; however, the mission of the hydrology community is not to quibble about what is or is not Bayesian but to develop effective methods for solving hydrological problems. In this sense, certainly, we do not see our two positions as being terribly far apart. It is true that in BARE, as implemented by TTGS, we have not explicitly differentiated between model (structural) error and data measurement errors. However, the implementation does distinguish between "model/parameter uncertainty" (dark shaded regions in the early portions of the hydrograph plots) and "other structure and data uncertainty" (light shaded regions). The estimated model is (models are) therefore not a deterministic, or true, representation of the system but has (have) been identified as having been assigned more than 99% of the probability mass when compared to the other members of the finite discrete approximation set of investigated models. As mentioned earlier, this should actually be interpreted as representing the neighboring "region of influence" of the identified model/parameter set. In contrast, although GLUE treats the sources of error implicitly without strong assumptions about a measurement error model, it does rely on assumptions about the suitable nature for a likelihood function and on ad hoc procedures for adjusting the likelihood function so that the prediction intervals bracket the observations. (In effect, GLUE might be viewed as ignoring data error and focusing primarily on model error.)

[22] Neither is it true that BARE is different from GLUE in admitting the possibility of equifinality of models. As mentioned in TTGS, and also above, the BARE approach clearly allows for multiple competing model structures and parameter sets and is therefore as equally wide ranging a methodology as GLUE is claimed to be. It is therefore hard to understand how BY can make the claim that the Bayesian methodology proposed in our paper is a special case of GLUE.

[23] Neither do we state anywhere that GLUE is subjective while BARE is objective. We do use the word subjective once, in the context of the GLUE methodology regarding the selection of the user-defined likelihood shape factor "N" and the threshold T that separates behavioral from nonbehavioral simulations. All identification methods are subjective to the extent that they rely on user-specified assumptions. In the theoretical development of BARE, the assumptions are that (1) the proposed set of model structures (and associated parameter uncertainty regions) are either suitable or adequate, (2) the prior model/parameter uncertainty can be stated in a reasonable manner, (3) there exists a (one or more) one-to-one and continuous invertible transformation such that the measurement errors in the transformed output space are mutually independent, each having an exponential power density of the form specified in TTGS (equation (9)), (4) the model parameters and log σ are independent, and (5) the recursive maximum likelihood estimate of σ is a suitable approximation. (Note that the theoretical development of BARE does not preclude the use of more than one different error model.) In the numerical implementation of BARE we add the implicit assumption that the set of all possible (suitable) model structures and parameter sets can be sufficiently approximated by a finite Monte Carlo sample of the model/parameter space. From a purely scientific point of view, all of these assumptions render the BARE procedure subjective until the assumptions have been properly tested. We do state that the GLUE methodology (as had been presented by Beven and coauthors at the time of writing of our paper) does not respect the rules of Bayesian inference, and briefly explain why. This is not to say that the formal GLUE procedure is not able to respect these rules when appropriate restrictions are placed on the choice of allowable likelihood functions [e.g., *Romanowicz et al.*, 1996].

[24] Having said all this, we agree with BY that the current developments of both BARE and GLUE are only partially adequate and that a preferable approach would be to separate out the effects of errors in the inputs, the errors in the model structures (including, we should add, parameter uncertainty), and real measurement errors in the outputs. However, as they point out, we do not currently know how to separate out these potential sources of error without making very strong and specific assumptions about their nature. This is the challenge that must be addressed in order to find a good way forward.

[25] In conclusion, we join Keith Beven and Peter Young in suggesting that neither BARE nor GLUE is a wholly satisfactory approach. We do believe that each highlights some important facets of the problem, while revealing areas that require considerable further attention. We hope that the discussions presented here and by *Beven and Young* [2003] will be received by the community in the spirit in which they are intended: that of healthy, mutually respectful and good natured debate. It is our mutual intention to raise interest in, and provoke discussion on, this important topic where there is no "right" or "wrong." Our hope is to motivate some bright and energetic young people to sink their teeth into these problems and progressively push the envelope of hydrologic modeling science further forward.

References

Beven, K., and P. Young, Comment on "Bayesian recursive parameter estimation for hydrologic models," by M. Thiemann et al., *Water Resour. Res.*, *39*, doi:10.1029/2001WR001183, in press, 2003.

Gupta, H. V., The devilish Dr. M, or, some comments on the identification of hydrologic models, paper presented at the 7th Annual British Hydrological Society National Symposium, Br. Hydrol. Soc., Newcastle-upon-Tyne, UK, Sept. 2000.

Romanowicz, R., K. J. Beven, and J. Tawn, Bayesian calibration of flood inundation models, in *Floodplain Processes*, edited by M. G. Anderson et al., pp. 333–360, John Wiley, New York, 1996.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian, Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535, 2001.

H. Gupta and S. Sorooshian, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ, 85721, USA. (hoshin@hwr.arizona.edu)

M. Thiemann, Riverside Technology, Inc., Fort Collins, CO 80525, USA.

M. Trosset, Department of Mathematics, College of William and Mary, Williamsburg, VA 23187, USA.